

ARESの関係モデルへの適用

APPLICATION OF ARES TO RELATIONAL MODELS

平川 正人

上林 憲行

市川 忠男

Masahito Hirakawa Noriyuki Kamibayashi

Tadao Ichikawa

広島大学

工学部

Faculty of Engineering, Hiroshima University

1. まえがき

近年、情報処理分野において取り扱われるデータ量の増加に関連して、大規模データベースの構成法、あるいは情報検索のための方式に大きな関心がよせられている。このようなデータベース技術の中でもデータ操作言語⁽¹⁾、特に非手続き的にデータの操作が行なえるもの⁽²⁾、データベースへのアクセスが簡単に行なえるもの⁽³⁾に注目が集まっている。ところでデータベースへ問合せを行なう場合、「20歳ぐらい」とか「黄色に近い」というようなあいまい性を伴った条件を表現したい事がしばしば起こる。この場合、これまでは取りうる値の範囲を不等式によって陽に表わさなければならず、ユーザにとって負担が大きい。それゆえ、上記のあいまい性を素直に受け入れる能力を持つことが効率の良いデータ操作言語であるための重要な条件であると考えられる。また、これに伴い表現上のあいまい性を許容し得るような処理システムが要求される。

筆者らは新しい連想方式に基づいた連想アクセス方式ARESを提案し、上述の問題に対する一つのアプローチを示した⁽⁴⁾。この連想方式では、データ間の関連の度合いを原データ上で測らず、誤り訂正処理を通して微細な特徴がぼやかされたデータに関して一致度の検定を行なう。これにより、データの関連度の評価を直接原データ上で行なう場合に要求される煩雑な計算処理は、誤り訂正処理と一致検定という単純な操作に置き換えられる。

本論文では、ARESを関係モデル方式のデータベースシステムにおけるインタラクティブな情報検索に適用する試みについて考察する。なお、ARESの連想方式は次のように概説される⁽⁴⁾。

(i) 個々のデータをデータ間の関連の度合いが符号間距離で表わされるように符号化する⁽⁵⁾。ここでデータの符号化はカテゴリ単位に行なう。
(ii) (i)によって符号化されたデータに誤り訂正処理を施し、類似したデータ間の共通の性質を導く。誤り訂正処理は使用する符号の符号長によって、いくつかのカテゴリをまとめたブロックを単位として行なう。
(iii) 内容アドレスリングによってデータ間の関連度評価を行なう。関連度はブロックの一致数によって評価する。

ARESを関係モデル方式のデータベースシステムに適用するにあたって、新しい比較演算子(≍)を導入した⁽⁶⁾。これによって、あいまい性を伴った問合せ(query)が可能となる。さらに、ARESでは連想サーチをより効果的に行なうために連想出カワード数の適応的制御機能が付与されている。つまり、連想出カワード数に制限が与えられているときには、総数がそのを超えない範囲で最大数の連想出カワードが求まる様に“制御”される。これはインタラクティブな検索において、適切なあいまい性の解釈を行なうために有効である。

2. 基本概念

2.1 将来のデータベースシステム

従来のデータベースシステムにおいては、問

合せは決定論的な記述によって行われる。このような決定論的な記述による問合せは、ユーザがデータベース内部の情報を持っていないので一般には不適当である。すなわち、ユーザは得ようとするデータが引きだされるまで機能的な問合せを繰り返し加さなければならぬ。繰り返しの問合せを行なうためユーザ、システムにとって共に負担が大きく、その上最終的に得られたデータが真にユーザの得ようとするデータであるかどうかは明らかでない。このような性格は今日のデータベースシステムにとって本質的な問題である。

これらの問題に対する改善の試みとして提案される新しい機能の幾つかを挙げれば次のとおりである。

- (1) 推論機構および学習機能を伴った知識データベース
- (2) コンピュータ支援システムを伴った問合せ
- (3) 多様なユーザズビューにダイナミックに適応する融通性のあるユーザインタフェース
- (4) 高レベルな検索機能を支援するアーキテクチャ

上記の点に関連して筆者らはARESを用い、あいまい性を伴った問合せが可能で、かつ融通性のあるユーザインタフェースを持った検索システムの構築を試みた。

2.2 新しい比較演算子の導入

これまでに多様な問合せ言語が提案され、実験されている。例えば写像型のSQL、例題型のQuery By Example (QBE) などがよく知られている⁽¹⁾。これらの言語で用意されている比較演算子には(=, ≠, ≥, >, ≤, <)があるが、いずれにもあいまい性を伴った問合せを可能にする比較演算子は用意されていない。我々は新しく“≐”という比較演算子を用意する。これは、例えば“AGE ≐ 20”という条件に対しては“年齢が20歳前後”の要素を、また“COLOR ≐ yellow”という条件に対しては“色が黄色に近い”要素を呼び出す比較演算子である。これにより、ユーザに

とってセマンティックギャップの少ない、より自然な問合せが可能となる。

図1に、「年齢が20歳前後の人の名前」を求める例を示す。ここで20歳前後という条件に対して5章で述べる制御を行なった結果、AGEという属性の要素が18から22までの値をとるレコードが選ばれれるものと仮定する。

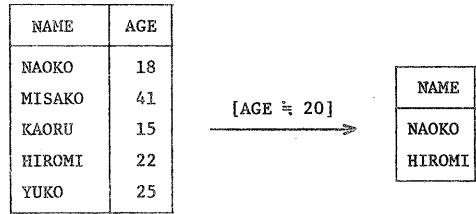


図1 新しい比較演算子の適用例

Fig.1 An example using new comparison operator

2.3 拡張された関係モデル

ここでは拡張された関係モデルについて述べることにする。従来の関係モデルにおいては定義域 (domain) D は要素の集合として定義されるが、個々の要素は他の要素との関連度に関するセマンティックな情報を一切含まない。我々は個々の要素が他の要素との関連度に関するセマンティックな情報を含むように定義域の定義を拡張し、これを D^* で表わす。図2にセマンティックな情報を含む要素の集合として定義される定義域 D^* の概念図を示す。

ある拡張された定義域 D_i^* に含まれる2つの要素 a_{ij}, a_{ik} 間の関連度を d_{jk}^i で表わす。この関連度 d_{jk}^i は、あいまい性を伴う問合せを行なう際に2つの要素間の関連度を評価する基本的な尺度となる。実際には、個々の要素を要素間の関連度が符号間距離に対応するように符号化する⁽²⁾。これによって関連度 d_{jk}^i は符号間距離にマッピングされる。

このような n 個の定義域 $D_1^*, D_2^*, \dots, D_n^*$ からなる直積 $D_1^* \times D_2^* \times \dots \times D_n^*$ の部分集合を関係 (relation) R と呼ぶ。ARESでは、関係 R の属性の各要素ごとに誤り訂正符号 (例えばハミング符号, リー符号) による誤り訂正処理を行なう。誤り訂正処理によって変換された要素 Q_{ij} を $c(Q_{ij})$ で表わす (図2の関係表

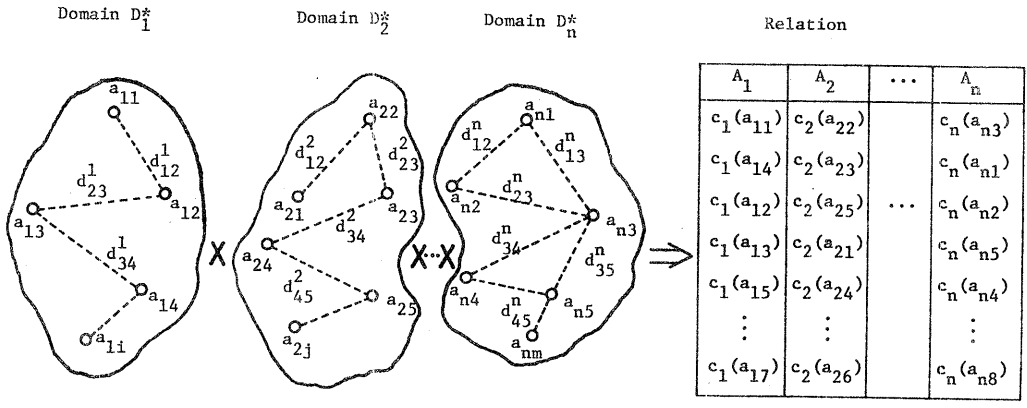


図2 拡張された関係モデル

Fig.2 Extended relational model

参照)。訂正可能な誤り距離 t の誤り訂正符号の使用を仮定すれば、2つの要素 Q_{ij} , Q_{ik} 間の関連度 d_{jk}^t が t 以下であるかどうかの評価は単に $C(Q_{ij})$, $C(Q_{ik})$ 上での一致検定に置き換えられる。

3. システム構成

図3にシステム構成を示す。システムは、(i) 入出力インターフェイス部 (Input/Output Interface Module), (ii) データベース制御部 (Database Management Module), (iii) ARES および関係データベース部 (ARES and Relational Database Module), の3つに分けられる。以下ではこれらの説明を行なう。

3.1 入出力インターフェイス部

ユーザはこの入出力インターフェイス部を通して会話形式でシステムとやりとりを行なう。その際、情報のやりとりはグラフィック機能を活用して行なう。

3.2 データベース制御部

データベース制御部は、入出力インターフェイス部と ARES および関係データベース部とのインターフェイスを受けもつとともに ARES の制御を行なう。

- (i) U I S (Users Interface Section)
U I S はユーザからのコマンドを解釈し、

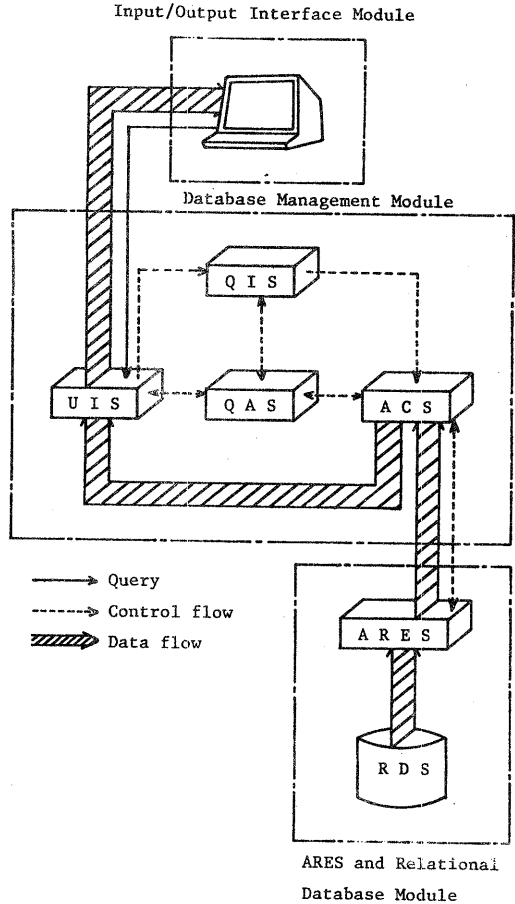


図3 システム構成

Fig.3 System organization

システムの処理結果を 入出力インターフェイス部に送る。

- (ii) QIS (Query Interpretation Section)
QIS は QBE ベースで記述された問合せを関係代数に翻訳する。
- (iii) QAS (Query Adaptation Section)
QAS はあいまい性を制御するためのパラメータを生成する。
- (iv) ACS (ARES Control Section)
ACS は ARES にコマンドおよび制御パラメータを送り, ARES の動作を制御する。

3.3 ARES および関係データベース部

(i) ARES

ARES は誤り訂正符号を用いた新しい連想方式により, 並列に高速連想サーチを行なうマルチマイクロプロセッサである。

(ii) RDS (Relational Database Store)

レコード (record) 間の関連の度合いが符号間距離で表わされるように符号化されている関係データベース

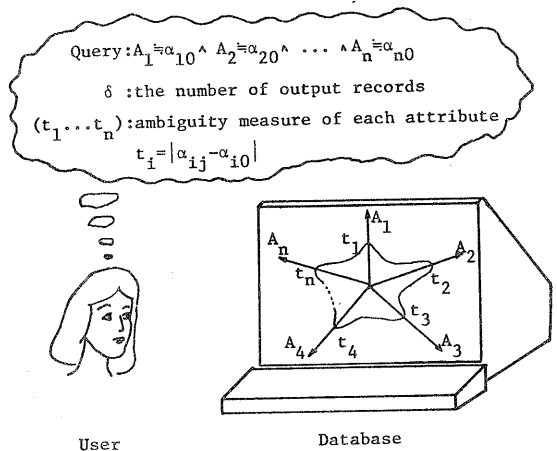
4. ユーザインターフェイス

4.1 データ操作機能

ここではユーザインターフェイスの観点からみたデータ操作機能について述べる。ユーザに提供されている機能には次のようなものがある。

- (i) ユーザは QBE ベースであいまい性を伴った問合せが行なえる。
- (ii) ユーザはあいまいさの程度を制御するために, 連想出力ワード数の制限 δ , および属性 (A_1, A_2, \dots, A_n) に対応するあいまい度 (ambiguity measure) (t_1, t_2, \dots, t_n) の指定が行なえる。

特にあいまい度 (t_1, t_2, \dots, t_n) の指定はグラフィック機能を用いて行なう (図4参照)。このようにして, ユーザは非手続的に問合せを行なうことが可能である。



User Database

図4 あいまい性を伴った問合せ

Fig.4 User's query with ambiguity

4.2 問合せの手順

本節ではあいまい性を伴う問合せを行なう際の手順について述べる。

- [Step 1] 端末からデータベーススキーマの名前および関係の名前を入力する。システムは QBE ベースモードに入り, 入力された名前 of 関係表を表示する。
- [Step 2] 図5のように QBE ベースで問合せを行なう。ここで "P." は「印字せよ」の略で, その属性を出力する。

CAR	CAR NAME	MAKER	PRICE [$\times 10^4$ yen]	TYPE
	P.280Z	P.NISSAN	≒ 120	

BODY COLOR	FUEL CONSUMPTION [km/l]	POWER [PS/rpm]	EXHAUST CAPACITY [cc]
≒ RED	≥ 15.0		≒ 1500

図5 問合せの記述例 (QBE)

Fig.5 An example of QBE base query

[Step 3] あいまい性を伴う問合せに対しては対応する属性にあいまい度を割り当てる。そして連想出力ワード数の制限 δ を図6のように入力する。

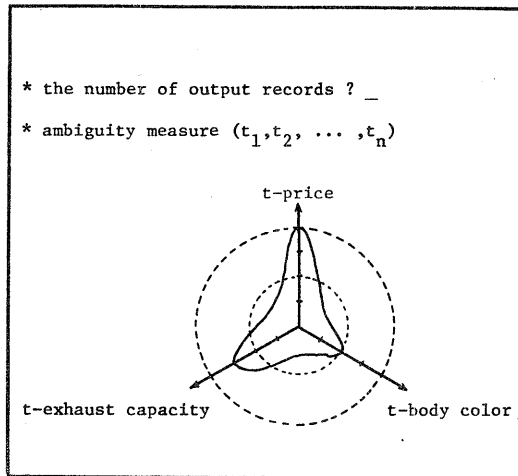


図6 あいまい度および δ の指定
Fig.6 Ambiguity assignment in computer assisted query

[Step 4] 問合せに従って処理が行なわれ、結果がユーザに戻される。

[Step 5] ユーザは結果をみて、満足の結果が得られるまで(i)問合せの条件、(ii) δ 、(iii)あいまい度、のいずれかを变更して処理を行なう。

図7に問合せを行なうための処理の流れ図を示す。

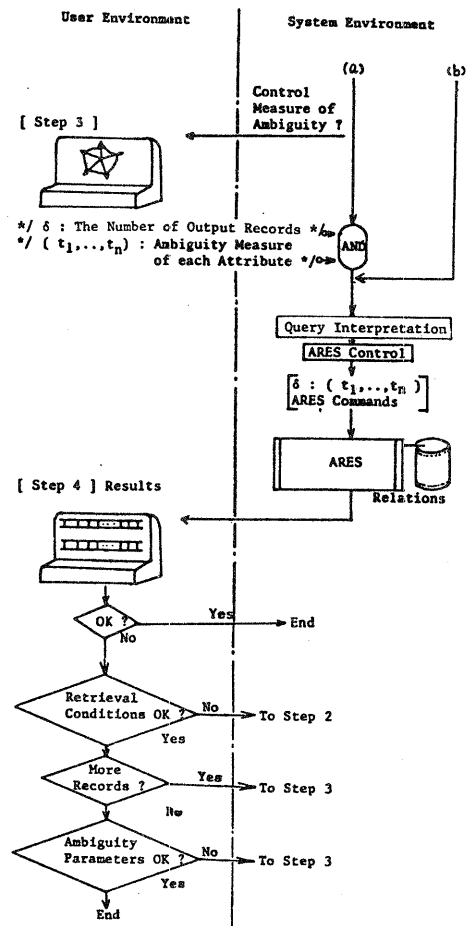
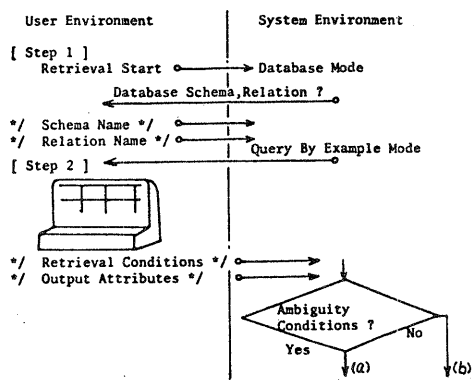


図7 問合せの処理流れ図

Fig.7 General flow for computer assisted query

5. 制御機構

ここではあいまい性を伴う問合せが、ARESの上でどのように実行されるかについて述べる。

5.1 諸定義

まず、あいまいさの評価を行なう機能が付加されている関係データベースシステムの構成に関する基本概念について述べる。

θ^* および θ は、それぞれ、セマンティックな関連度が符号間距離で表現されている要素からなる集合のクラス、あるいは関連度づけがなされていない要素からなる集合のクラスを表わすものとする。ここで、一般のデータベース

では要素間の関連度が符号間距離に対応するよ
うな符号の割り当ては、すべてのユーザに対
して不遷性があるように行われなければならない。

$D_i \in (\mathcal{D} \cup \mathcal{D}^*)$, $1 \leq i \leq n$, を満たす n 個の
集合 D_1, D_2, \dots, D_n を考える。これらの直積
 $D_1 \times D_2 \times \dots \times D_n$ の部分集合を関係記と表わす。
集合 D_1, D_2, \dots, D_n を定義域と呼び、同一の集
合の重複を許す。関係記における i 番目のレ
コードを R_i で表わし、 $D_i \in \mathcal{D}$ であるような定
義域 D_i に対応する属性 (attribute) A_i の集合を
 \mathcal{A} , $D_j \in \mathcal{D}^*$ であるような定義域 D_j に対応する
属性 A_j の集合を \mathcal{A}^* とする。またレコード R_i
における属性 A_j の要素を R_{ij} で表わす。

さらに属性の条件を示す " $A_j = \alpha$ " ($A_j \in \mathcal{A}^*$,
 $\alpha \in D_j \in \mathcal{D}^*$) を A タイプ質問項 (A-type query),
" $A_i = \beta$ " ($A_i \in \mathcal{A}$, $\beta \in D_i \in \mathcal{D}$) を B タイプ質問
項 (B-type query) と呼び $\mathcal{Q}_A, \mathcal{Q}_B$ で表わす。
両者を区別する必要がない場合には、単に質問
項と表わす。

実際には、ユーザは図5で示したようにQBE
ベースで問合せを行なう。"280 匹" のよう
に下線を引いたものを例要素といい、「たとえば
280 匹を」を意味する。" \approx RED", " ≥ 15.0 "
等は問合せの条件を示す定数要素であり、これ
によって A タイプ質問項, B タイプ質問項を表
現する。すなわちシステムは表形式で表現さ
れた問合せに従い、理想出力ワード数の制限 δ
の範囲内で最大数のデータを検索する。

5.2 関連度の評価

それぞれの属性について、属性別に要素間の
関連度が、各要素に割り当てられた符号の符号
間距離に対応するように符号化されていると仮
定する。

A タイプ質問項 $A_j = \alpha$ においては、 α と R_{ij}
(i 番目のレコード R_i における属性 A_j の要素)
にそれぞれ割り当てられた符号の符号間距離が
 d_j (属性 A_j に対するあつまり度) 以下であるよ
うなレコード R_i が (距離 d_j で) 選ばれる。
B タイプ質問項 $A_i = \beta$ においては、 β と R_{ij} に
それぞれ割り当てられた符号の符号間距離が 0
であるようなレコード R_i が (距離 0 で) 選
ばれる。すなわち、あつまり度 d_j は A タイプ
質問項を満たすレコードを選べる際のあま
りさの程度を規定する。

5.3 θ 制御

x 個の質問項からなる $Q = \mathcal{Q}_1 \wedge \mathcal{Q}_2 \wedge \dots \wedge \mathcal{Q}_x$ を
考える。ここで $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_x$ は A タイプ質問
項あるいは B タイプ質問項である。もし \mathcal{Q}_j
($1 \leq j \leq x$) の少なくとも一つが A タイプ質問項
であれば、4.2 の問合せ手順に従って理想出力ワ
ード数の制限 δ およびあつまり度をを入力する。
すべての \mathcal{Q}_j ($1 \leq j \leq x$) が B タイプ質問項であ
れば、従来の関係モデルと等価となる。

i 番目のレコード R_i に対して \mathcal{Q}_j ($1 \leq j \leq x$) を
評価し、満足する質問項の数が θ (θ はある正
整数とする) 以上であるようなレコードを選
び出す。選んだレコード数が理想出力ワ
ード数に対する制限 δ を超えた場合には θ の値
を 1 つ増やし、 δ を満たない場合には θ の値を
1 つ減らして再度 Q に関する評価を行なう。
このようにして δ 以内で、かつ最大数のレ
コードが選ばれるように θ の値をヒューリス
ティックに制御する。アルゴリズムに記述したこ
の θ 制御のアルゴリズムを付録に示す。

次に、例を用いて θ 制御の説明を行なう。
付表 1 に示す CAR という関係表に対して、あ
つまり性を伴った問合せを行なうことは可能。
4.2 で述べた手順に従い、図8のような問合せ
を行なう。

CAR	CAR NAME	MAKER	PRICE [$\times 10^4$ yen]	TYPE
	P.CIVIC		P. ≈ 11.0	P.HATCH BACK

BODY COLOR	FUEL CONSUMP- TION [km/l]	POWER [PS/rpm]	EXHAUST CAPACITY [cc]
	P. ≥ 14.0		P. ≈ 1500

図8 問合せ例

Fig.8 An example of query

また、図6に示す形式で理想出力ワード数の
制限として $\delta = 5$, PRICE, EXHAUST CAPACITY
の2つの属性に対してあつまり度を、 d_j をそれ

CAR NAME	PRICE [$\times 10^4$ yen]	TYPE	FUEL CONSUMPTION [km/l]	EXHAUST CAPACITY [cc]
CORSA	96	HATCH BACK	15.5	1500
MIRAGE	105	HATCH BACK	14.5	1600
FAMILIA	107	SALOON	16.5	1500
CIVIC	94	HATCH BACK	15.0	1500
SPRINTER	109	HARD TOP	15.0	1500

図9 検索結果

Fig.9 A final answer

ぞい入力したとする。

ここでAタイプ質問項 "PRICE ≥ 110 " に対しては100から120までの要素が, "EXHAUST CAPACITY ≥ 1500 " に対しては1400から1600までの要素が選ばれると仮定する。とりうる要素の範囲はあいまい度 θ , および各要素に割り当てられた符号に依存する。

$\theta = 1$ で θ 制御を行なうと4つの質問項 "PRICE ≥ 110 ", "TYPE = HATCH BACK", "FUEL CONSUMPTION ≥ 14.0 ", "EXHAUST CAPACITY ≥ 1500 " のうち, 少なくとも一つの質問項を満たすレコードが選ばれる。CARという関係においては16個のレコードがこれに該当し, 出力数の制限 $\delta = 5$ を超えている。それゆえ θ の値を1だけ増やして再度 θ 制御を行なう。以降, θ の値をヒューリスティックに適應させながら, δ を超えない範囲で最大数のレコードが選ばれるまで θ 制御を繰り返す。

最終的に $\theta = 3$ で5個のレコードが選ばれ, 図9のような結果がユーザに示される。

6. おわりに

連想プロセッサARESは, あいまい性を伴う処理に適した新しい連想方式に基づいて大量情報の高速サーチを行なう。この処理過程へのあいまい性の導入に加えて, 連想出力ワード数の可制御性がARESの特徴となっている。

本論文では, 関係データベース環境への応用に対して, あいまい性を伴った問合せに対する検索の制御機構を明らかにするとともに, インタラクティブな情報検索環境の下でARESが

有効に働くことを示した。なお θ 制御は, 取り扱う関係が1NFのみで, 関係の間の演算が行なわれない範囲で考えている。またシステムをシミュレートしたものがHP-1000上で動いており, 今後ユーザの好みなど, 主観的要素をとり入れたより実際的な環境設定のもとでの実験を進める予定である。なお, ARESは画像データベースの検索に関するシステムの構成にも有効である⁽⁷⁾。

終りに, 本研究の進展過程で積極的な御協力を頂いた本学助教授野野原 誠, ならびに, 有益な御意見を頂いた徳山高専 岡野 博二助教授に感謝する。

文献

- (1) Date, C. J. : "An Introduction to Data Base Systems", Addison-Wesley, 2nd Edition (1977).
- (2) Astrahan, M. M. and Chamberlin, D. D. : "Implementation of a Structured English Query Language", Commun. ACM, 18, 10, pp.580-588 (Oct. 1975).
- (3) Chang, N. S. and Fu, K. S. : "A Relational Database System for Images", TR-EE 79-28, Purdue University (May 1979).
- (4) 市川, 坂村, 諸隈, 相磯 : "連想プロセッサARES", 信学論誌(D), J61-D, 10, pp.743-750 (1978-08).
- (5) 榎本, 市川 : "線形グラフの二元符号空間への整合方法", 信学論誌(C), 57-C, 8, pp.375-382 (1968-08).
- (6) Ichikawa, T., Kikuno, T., Kamibayashi, N. and Hirakawa, M. : "ARES in Relational Database Environments", Proc. COMPSAC '80, pp.557-561 (Oct. 1980).
- (7) Ichikawa, T., Kikuno, T. and Hirakawa, M. :

"A Query Manipulation System for Image Data Retrieval by ARES", Proc. 2nd IEEE Workshop on Picture Data Description and Management", pp.61-67 (Aug. 1980).

付録

θ 制御のアルゴリズムを以下に示す。

```

procedure THETA_CONTROL(Q, R,  $\delta$ , t)
comment associate as large a number of
    records in R as possible within
    the limit  $\delta$ 

begin
     $\theta := \theta_0$ ;
    sw:=0;
INIT: hit_number:=0;
    clear Buffer;
    R1:=first record in relation R;
CELL: if Q-RELEVANCY(R1, Q, t,  $\theta$ )=true then
    begin
        hit_number:=hit_number+1;
        put R1 into Buffer;
        if hit_number>> $\delta$  then begin
            sw:=1;
             $\theta := \theta + 1$ ;
            goto INIT;
        end;
    fi
    end;

    fi
end;

fi
if all records in relation R have
    already been processed then
    begin
        if hit_number< $\delta$  then
            begin
                if sw=1 then goto END;
                else begin
                     $\theta := \theta - 1$ ;
                    goto INIT;
                end;
            end;
        else goto END;
    end;

    fi
    else begin
        R1:=next record in R;
        goto CELL;
    end;

```

END: end of algorithm

```

end;
end THETA_CONTROL

```

```

procedure Q-RELEVANCY(Ri, Q, t,  $\theta$ )
comment relevancy estimation between a query
    Q=q1  $\wedge$  q2  $\wedge$  ...  $\wedge$  qx and a record Ri;
    let qj is the A-type query (Aj =  $\alpha_j$ )
    or the B-type query (Aj =  $\beta_j$ )

begin
    count:=0;
    repeat j=1 until x
        Rij:=value of attribute Aj of Ri;
        case: qj is the A-type query :
            if a code distance between Rij and
                 $\alpha_j$  is less than or equal to tj
            then count:=count+1;
        : qj is the B-type query :
            if a code distance between Rij and
                 $\beta_j$  is less than or equal to 0
            then count:=count+1;
        esac
    taepr
    if count>> $\theta$  then Q-RELEVANCY(Ri, Q, t,  $\theta$ )=true;
    else Q-RELEVANCY(Ri, Q, t,  $\theta$ )=false;
    fi
end;
end Q-RELEVANCY

```

なお、 θ の適応的制御過程を図 10 に示す。
ここで θ^* は θ の最終値を表わしている。

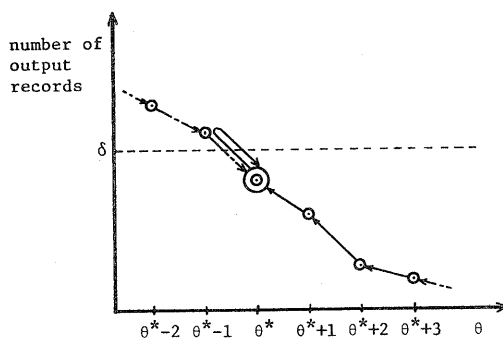


図 10 θ の適応的制御過程

Fig.10 Process of θ -control

付表1 サンプルデータ

CAR	CAR NAME	MAKER	PRICE [$\times 10^4$ yen]	TYPE	BODY COLOR	FUEL CONSUMP- TION [km/l]	POWER [PS/rpm]	EXHAUST CAPACITY [cc]
	CORONA	TOYOTA	156	SEDAN	BROWN	10.0	135	2000
	CIVIC	HONDA	98	HATCH BACK	WHITE	16.0	72	1300
	117 COUPE	ISUZU	220	COUPE	RED	9.0	120	2000
	FAMILIA	MAZDA	85	HATCH BACK	WHITE	17.0	74	1300
	CORSA	TOYOTA	96	HATCH BACK	RED	15.5	83	1500
	CAPELLA	MAZDA	136	HARD TOP	WHITE	12.0	110	2000
	ACCORD	HONDA	145	HATCH BACK	BROWN	12.5	95	1800
	COROLLA	TOYOTA	137	HARD TOP	YELLOW	12.5	115	1600
	MIRAGE	MITSUBISHI	105	HATCH BACK	YELLOW	14.5	88	1600
	COROLLA	TOYOTA	97	LIFT BACK	BLUE	15.0	72	1300
	FAMILIA	MAZDA	107	SALOON	RED	16.5	85	1500
	CIVIC	HONDA	94	HATCH BACK	RED	15.0	85	1500
	GEMINI	ISUZU	164	SEDAN	BLACK	11.5	130	1800
	CELICA XX	TOYOTA	200	HATCH BACK	WHITE	8.7	145	2800
	SPRINTER	TOYOTA	109	HARD TOP	WHITE	15.0	80	1500
	LANCER EX	MITSUBISHI	109	SEDAN	SILVER	13.5	86	1600
	LEONE	SUBARU	115	SEDAN	BLUE	13.5	87	1600
	FAMILIA	MAZDA	97	HATCH BACK	WHITE	17.0	74	1300
	SILVIA	NISSAN	175	HATCH BACK	RED	11.0	120	2000
	SUNNY	NISSAN	112	SEDAN	BLUE	15.0	74	1300